УДК 519.245:519.688

ПРИМЕНЕНИЕ АДАПТИВНОГО БИНОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ В МЕТОДЕ ПОИСКА ГЛОБАЛЬНОГО МИНИМУМА SIMULATED ANNEALING

А.А. Хамухин

Томский политехнический университет
E-mail: alex@ad.cctpu.edu.ru

В методе поиска глобального минимума Simulated Annealing предложено использовать бинормальное распределение плотности вероятности следующего шага, моды которого смещены относительно текущего локального минимума, а расстояние между ними и дисперсии функционально связаны с искусственной температурой. Показана эффективность реализаций подхода с помощью численных расчетов.

Метод поиска глобального минимума, известный в литературе под названием Simulated Annealing (SA), или метод «имитации отжига» применяется при построении математических моделей и решения сложных оптимизационных задач в нейрокомпьютерной технике, нефтегазогеологии, микроэлектронике, микробиологии, ядерной физике и др. [1–5]. Метод SA, предложенный Киркпатриком в 1982 г. [6], имеет различные алгоритмические реализации, некоторые из которых включены в такие известные вычислительные пакеты, как Mathematica NMinimize, STATISTICA Neural Networks и ряд других. Главным его достоинством является теоретическое доказательство сходимости к глобальному минимуму при использовании распределения Больцмана [1]. Однако практический сравнительный анализ методов глобальной оптимизации показал, что метод SA является наиболее «хрупким», т. е. заметно зависит от выбранных параметров поиска и нуждается в дополнительной настройке [6]. Так, например, в работе [7] утверждается, что не существует универсального по эффективности алгоритма для разных задач глобальной оптимизации и задача их разработки и модификации остается открытой.

Целью работы автора является создание эффективных инструментальных средств настройки па-

раметров метода SA, а в данной статье публикуются первые результаты модификации алгоритма, направленной на повышение его эффективности. Поставленная цель достигается путем уточнения аналогии оптимизационного метода SA с естественным процессом, послужившим основанием для его разработки.

Метод SA базируется на аналогии с физическим процессом кристаллизации металла при постепенном переходе из жидкой фазы в твердую, в результате которого достигается глобальный минимум энергетического состояния (энтропии). Как известно, при стремлении температуры любой равновесной термодинамической системы к абсолютному нулю ее энтропия стремится тоже к нулю (третье начало термодинамики). Однако аморфные тела не относятся к равновесным системам, поэтому их конечная энтропия при понижении температуры зависит от того, как производится сам процесс охлаждения. В природе и в человеческой деятельности есть образцы такого «удачного» понижения энтропии, например: кристаллизация атомов углерода при образовании алмаза, секреты дамасской стали. Хотя из третьего начала термодинамики следует недостижимость температуры, равной абсолютному нулю, для практического охлаждения термодинамической системы до некоторой минимальной температуры рекомендуется чередовать изотермическое сжатие и адиабатическое расширение. При первом процессе происходит отвод теплоты, а при втором – уменьшение температуры системы.

Особое значение в этом процессе имеет скорость охлаждения, поскольку из теории кристаллизации известно, что если металл очень сильно переохладить, то число центров и скорость роста кристаллов равны нулю, жидкость не кристаллизуется, образуется аморфное тело. Продолжая эту аналогию на молекулярно-атомный уровень, следует отметить, что атомы имеют дискретный энергетический спектр, т. е. каждый атом может принимать только определенные энергетические состояния, ограниченные потенциальными барьерами. Поэтому быстрая потеря его энергии может привести к тому, что он останется в более высоком энергетическом состоянии из-за невозможности преодолеть потенциальный барьер. Распределение энергетических уровней описывается известным соотношением:

$$P(e) = A\exp\left(\frac{-e}{kT}\right), \quad (1)$$

где $P(e)$ – вероятность того, что система находится в состоянии с энергией e, k – постоянная Больцмана, T – температура по шкале Кельвина, A – нормировочный коэффициент.

Таким образом, по мере снижения температуры вероятность высокоэнергетических состояний снижается, и при приближении температуры к нулевой достигается глобальный минимум энергетического состояния.

В терминах оптимизационного алгоритма это означает обеспечение выхода из текущей ловушки локального минимума, и не возвращение в предыдущую ловушку. Для этого, как минимум, необходимо разрешать движение даже в том направлении, где целевая функция возрастает, но с некоторыми ограничениями. На практике это реализуется с помощью распределения вероятности принятия следующего шага, аналогичной приведенному выше распределению Больцмана (1):

$$P_{ijk} = \exp\left(\frac{f(x_i) - f(x_j)}{T_k}\right), \quad (2)$$

где P_{ijk} – вероятность перехода при данном T_k из i-й точки пространства поиска в j-ю, если значение целевой функции f возросло, T_k – искусственная температура, которую необходимо снижать на k-м шаге.

Каждый пробный шаг по пространству координат поиска генерируется случайно по некоторому закону (Больцмана, Коши, Гаусса) [1]. Однако эти алгоритмы обладают очень низкой скоростью сходимости и для практического применения требуется их модификация.

Так, естественным выглядит аналогия оптимизационного алгоритма и процесса кристаллизации. На наш взгляд, каждый локальный минимум целевой функции следует рассматривать как отдельный атом. Область в пространстве координат вокруг этого минимума, где целевая функция не убывает, можно считать зоной действия, потенциальной ямой или «ловушкой» локального минимума, а расстояние от центра до границы зоны можно соотнести с местоположением потенциального барьера атома. Тогда размер следующего шага, генерируемого алгоритмом, целесообразно связать аналогией с кинетической энергией атома, которая в свою очередь связана с температурой по соотношению (1) или с искусственной температурой по соотношению (2). В этом случае порядок снижения температуры должен быть таким, чтобы каждый атом имел достаточно энергии для преодоления потенциального барьера и перехода в более низкоэнергетическое состояние, но недостаточно энергии для возврата в более высокоэнергетическое состояние, что и обеспечит выход из текущей ловушки локального минимума.

В настоящей работе исследовано на примерах и предложено:

1. Генерировать пробный шаг по пространству координат поиска с использованием бинормального распределения плотности вероятности (два распределения Гаусса, моды которых смещены влево и вправо относительно текущего локального минимума).
2. Расстояние между модами бинормального распределения и их дисперсии сделать функциями искусственной температуры. При понижении температуры эти центры должны сближаться, а дисперсии уменьшаться. В итоге центры

будут сливаться в одно нормальное распределение в области, определенной как зона глобального минимума.

Двухмодальность бинормального распределения, центр которого расположен в текущем локальном минимуме, даст для самого процесса поиска несколько преимуществ. *Во-первых*, наиболее вероятный шаг будет смещен от точки текущего локального минимума (влево и вправо), что более соответствует приведенной выше физической аналогии. Это должно обеспечивать более быстрый выход из ловушки текущего локального минимума и переход в зону другого, т. к. за счет двухмодальности снижена вероятность шагов вблизи локального минимума, которые фактически являются излишними, если текущий минимум не окажется глобальным. *Во-вторых*, функциональная зависимость расстояния между модами от искусственной температуры также более соответствует приведенной выше физической аналогии и позволит учитывать вид самой целевой функции.

В математической форме вид предлагаемого бинормального распределения представлен ниже:

$$P(X_j) = \frac{1}{S_1\sqrt{2\pi}} \exp\left(-\frac{(X_j - M_1)^2}{2S_1^2}\right), \text{если } X_j < X_i; \quad (3)$$

$$P(X_j) = \frac{1}{S_2\sqrt{2\pi}} \exp\left(-\frac{(X_j - M_2)^2}{2S_2^2}\right), \text{если } X_j > X_i; \quad (4)$$

$$S_1 = F_1(T_k), \quad S_2 = F_2(T_k), \quad (5)$$

$$M_1 = X_i - d_1 T_k, \quad M_2 = X_i + d_2 T_k, \quad (6)$$

$$T_0 > T_1 > T_2 > ... > T_k > ... > T_z = 0, \quad (7)$$

где X_i – текущая точка в пространстве координат поиска, в которой находится центр бинормального распределения плотности вероятности следующего шага $P(X_j)$; X_j – следующая точка, в которую с некоторой вероятностью будет перемещен центр бинормального распределения; M_1, M_2 – математические ожидания; S_1, S_2 – дисперсии нормальных распределений, составляющих бинормальное, которые являются некоторыми функциями F_1, F_2 от искусственной температуры T_k; d_1, d_2 – коэффициенты смещения мод бинормального распределения относительно центра.

В случае возрастания значения целевой функции после генерации X_j вопрос о переходе в нее из точки X_i решается по-прежнему с помощью распределения (2). Наличие функциональных зависимостей (5), (6) параметров распределения плотности вероятности следующего шага от искусственной температуры позволяют говорить о его адаптивности по ходу поиска.

На рис. 1 представлен примерный графический вид адаптивного бинормального распределения по одной координате в начале (*a*), середине (*b*) и в конце поиска (*c*) на примере одной из тестовых целевых функций.

Бинормальное распределение легко обобщается и на многокоординатный поиск, например, для двух координат его примерный вид представлен на рис. 2.

Таким образом, адаптивность плотности распределения вероятности следующего шага обеспечивается предварительной настройкой и функциональным изменением по ходу поиска следующих параметров: искусственная температура, расстояние между модами бинормального распределения, дисперсии нормальных распределений, составляющих бинормальное по каждой координате.

Для исследования предлагаемого алгоритма и сравнения его с иными реализациями метода SA (с равномерным и нормальным распределениями) в математическое описание (3)–(7) были внесены условия симметрии бинормального распределения, прямой пропорциональности дисперсий искусственной температуре и пропорционального снижения самой искусственной температуры через заданное число шагов:

$$d_1 = d_2 = d_0, S_1 = S_2 = s_0 T_k, T_{k+1} = t_0 T_k, 0 < t_0 < 1, \quad (8)$$

где d_0, s_0, t_0 – некоторые константы для одного цикла поиска, которые являются настроечными параметрами алгоритма и зависят от вида целевой функции.

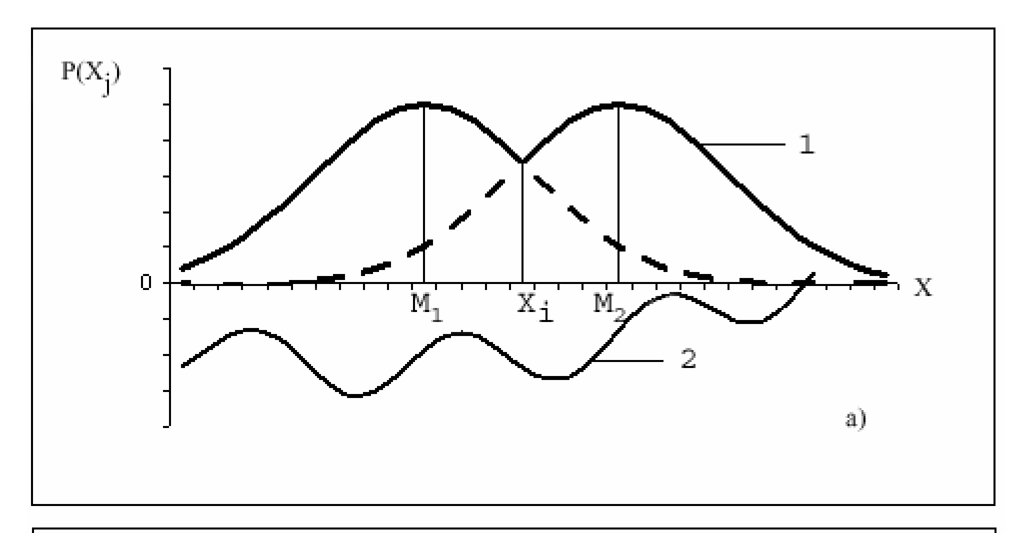
Для реализации алгоритма (2)–(8) была написана специальная программа пошагового выполнения и наблюдения за ходом поиска с помощью графиков и таблиц [8]. Наблюдение и сравнение трех реализаций метода SA выполнялось на нескольких тестовых целевых функциях вида:

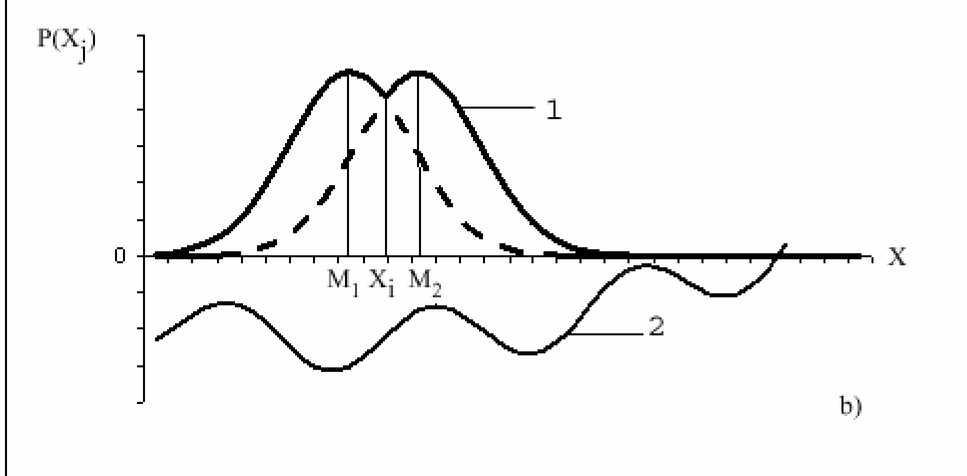
$$f(x) = \frac{x^2 + a\sin bx}{c}.$$

Эти функции имеют бесконечное множество локальных минимумов, несколько из которых достаточно близки и являются ловушками для любого алгоритма поиска глобального минимума. Численные исследования проводились для различных сочетаний коэффициентов a,b,c и показали примерно подобные результаты. В качестве характерного примера в таблице приведены результаты для набора коэффициентов a=1, b=5, c=5. Тестовая функция при этом имеет глобальный минимум в точке x=–0,29083932 с точностью до 10^{-8}.

Очевидно, что зависимость результатов поиска от начальной искусственной температуры остается для всех рассмотренных реализаций метода. Однако при использовании бинормального распределения вероятность попадания в ловушку локального минимума ниже, чем для нормального распределения, алгоритм реализации которого при начальной искусственной температуре 200 попадал и не выбирался из ловушки локального минимума.

Из таблицы видно, что алгоритм реализации метода SA с равномерным распределением требует значительно большего количества шагов, чем при реализации методов, основанных на нормальных





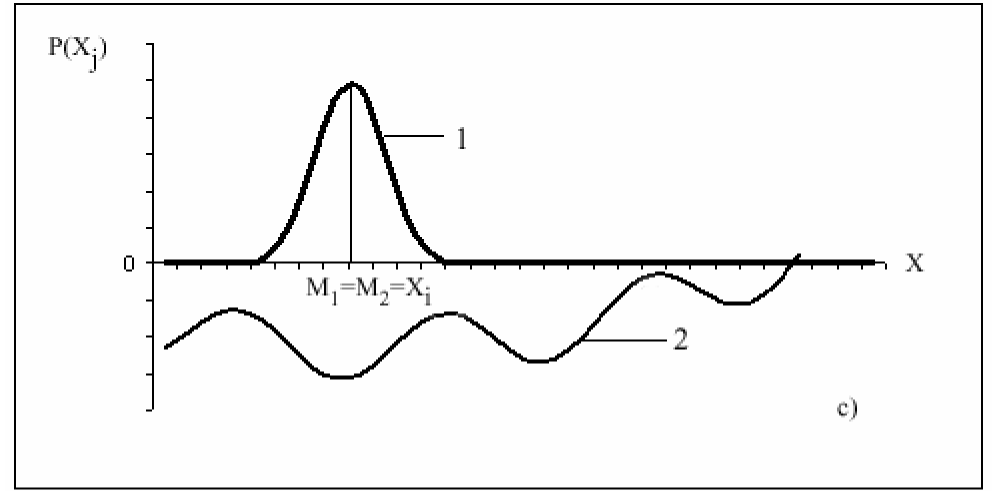


Рис. 1. *Пример адаптивного бинормального распределения плотности вероятности следующего шага (1) при поиске глобального минимума целевой функции (2)*

распределениях для обеспечения сопоставимой с ними погрешности поиска. Преимущество бинормального распределения по сравнению с нормальным при поиске глобального минимума заметно проявляется при больших количествах заданных шагов (например, при начальной искусственной температуре 800 для заданной выше тестовой целевой функции). Это объясняется значительным снижением количества вычислений значений целевой функции для генерируемых шагов вблизи текущих локальных минимумов, которые, по сути, являются холостыми шагами. Поэтому, чем больше

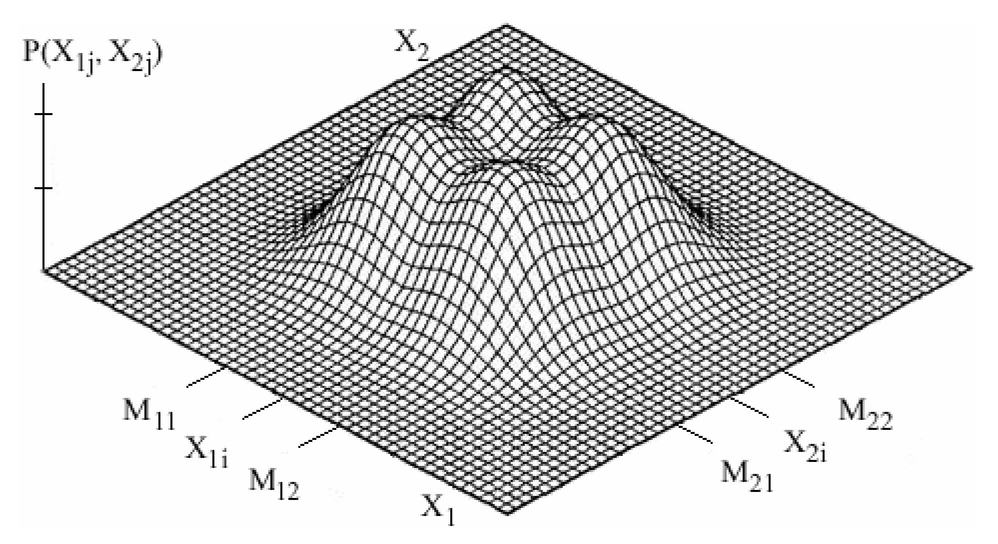


Рис. 2. *Пример двухмерного (координаты X_1,X_2) бинормального распределения плотности вероятности следующего шага $P(X_{1,j},X_{2,j})$*

Таблица. *Результаты поиска глобального минимума методом SA с различными распределениями плотности вероятности следующего шага*

Распределение	Начальная "температура"	Количество шагов	Найденная точка минимума	Относительная погрешность, %
Равномерное	400	65	−0,25524	12,24
	1800	402	−0,27776	4,49
Нормальное	200	35	−1,45146	399,05*
	400	65	−0,28331	2,59
	800	130	−0,29512	1,47
Бинормальное	200	35	−0,27981	3,79
	400	65	−0,28442	2,21
	800	130	−0,29139	0,19

** ловушка локального минимума*

локальных минимумов будет иметь целевая функция в области поиска, тем больше будет выигрыш в скорости поиска (при заданной погрешности) или в точности поиска (при заданном количестве шагов) предлагаемого алгоритма реализации метода SA с адаптивным бинормальным распределением.

Выводы

1. Показана возможность повышения эффективности поиска глобального минимума целевых функций методом Simulated Annealing за счет применения адаптивного бинормального распределения плотности вероятности следующего шага поиска по сравнению с нормальным и равномерным распределениями.
2. Показано, что предложенный подход снижает погрешность поиска при заданном количестве шагов или уменьшает их количество при сопоставимой погрешности за счет уменьшения количества вычислений целевой функции вблизи локальных минимумов.
3. Расчетами установлено, что за счет специфики формы предложенного распределения наблюдается снижение количеств попаданий алгоритма поиска в ловушки локальных минимумов по сравнению с нормальным распределением плотности вероятности следующего шага.

СПИСОК ЛИТЕРАТУРЫ

1. Уоссерман Ф. Нейрокомпьютерная техника: Теория и практика. Пер. с англ. – М.: Мир, 1992. – 240 с.
2. Хамухин А.А. Математическое моделирование процесса сепарации пластовой нефти // В сб.: Математическое и программное обеспечение САПР. Под ред. В.К. Погребного. – Томск, 1997. – Вып. 1. – С. 167–172.
3. Глебов А.Л., Лопатников С.Ю. Новый алгоритм моделируемого отжига для оптимизации КМОП-схем // Информационные технологии. – 1998. – № 1. – С. 25–28.
4. Grigoriev I.V., Rakhmaninova A.B., Mironov A.A. Simulated annealing for alpha-helical protein folding: searches in vicinity of the «molten globule» state // J. Biomol. Struct. Dyn. – 1998. – V. 16(1). – P. 115–122.
5. Ali A.M. Use of Monte Carlo Method with Simulated Annealing Algorithm in the Analysis of Mossbauer Spectra // Fizika B. – 2002. – V. 11. – № 1–4. – P. 19–26.
6. Минаков И.А. Сравнительный анализ некоторых методов случайного поиска и оптимизации // Известия Самарского научного центра РАН. – 1999. – № 2. – С. 286–293.
7. Орлянская И.В. Современные подходы к построению методов глобальной оптимизации // Электронный журнал «Исследовано в России» (http://zhurnal.ape.relarn.ru/articles/2002/189.pdf). – 2002. – Т. 5. – С. 2097–2108.
8. Хамухин А.А. http://ad.cctpu.edu.ru/Personal/Alex/Zhtml/EXAMP/SimAnn51.xls